Теория сигналов и систем

УДК 534.78

К.П. Пилипенко, А.Н. Продеус, канд. техн. наук

Использование кумулянтных коэффициентов для определения пола диктора

Рассмотрена возможность использования кумулянтных коэффициентов в качестве классификационных признаков для определения пола диктора. Построены классификаторы с помощью логистической регрессии и линейного разделения.

The possibility of using the cumulant coefficients as classification criteria in determining the gender is considered. Classifiers were constructed by using logistic regression and linear separation.

Ключевые слова: *плотность вероятности сигнала, октавная полоса частот, кумулянтные коэффициенты, тестирование.*

Введение

Определение пола диктора является одной из задач голосовой биометрии [1, 2]. В настоящее время идентификация пола диктора используется в системах обеспечения безопасности. Кроме того, определение пола диктора позволяет проводить более точную настройку систем распознавания речи.

Решение данной задачи принципиально не отличается от решения любой другой задачи автоматического распознавания образов и состоит из стандартных этапов:

– формирование обучающей базы данных;
– выделение классификационных признаков;
– выбор и обучение модели;
– построение решающего правила (классификатора).

Одним из наиболее важных этапов, который, в конечном счете, будет определять качество классификации, является выбор классификационных признаков.

Как правило, в качестве информационного параметра, по которому проводится идентификация пола диктора, используют частоту основного тона. Однако, как показывает практика, одной частоты тона недостаточно для достоверной классификации пола диктора, поэтому вектор признаков дополняют кепстральными параметрами [1, 2].

Цель данной работы состоит в том, чтобы показать принципиальную возможность использования кумулянтных коэффициентов в качестве классификационных признаков при определении пола диктора.

1. Выбор классификационных признаков

В работе [3] исследовались законы распределения мгновенных значений речевых сигналов для мужских и женских голосов. Результаты исследований показали, что законы распределения мужских и женских голосов наиболее различны в крайних октавных полосах частот – около 125 и 8000 Гц. Рассмотрим речевой сигнал в октавной полосе частот со среднегеометрической частотой 125 Гц. На рис. 1 изображены графики оценок плотностей вероятностей мгновенных значений сигнала, полученных для фонограмм мужского и женского голосов в указанной полосе частот.

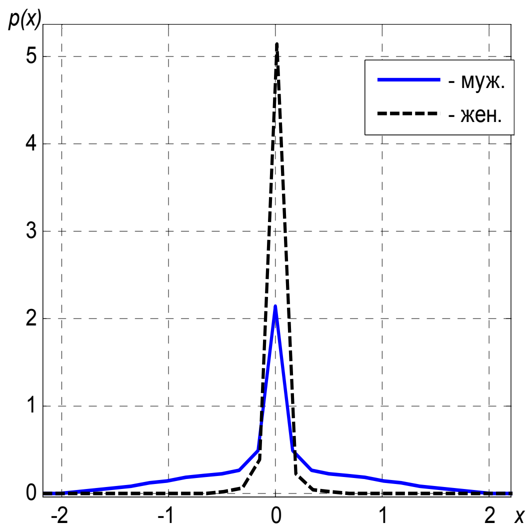

Рис. 1

Как следует из рис. 1 плотности вероятностей для мужских и женских речевых сигналов существенно отличаются. Данный факт может стать основой для классификации пола диктора.

Следует отметить, что указанные плотности вероятностей могут быть аппроксимированы смесью нормальных распределений. Частный случай смеси нормальных распределений был предложен в качестве модели плотности вероятностей речевого сигнала в работе [4]. Задача аппроксимации плотности вероятностей одновершинной смесью распределений рассмотрена в работе [5]. В работах [5, 6] показано, что плотность вероятностей одновершинной гауссовской смеси однозначно определяется кумулянтными коэффициентами γ_4 и γ_6.

Таким образом, если в качестве модели плотности вероятностей речевого сигнала принять одновершинную двухкомпонентную гауссовскую смесь [5], то для идентификации распределения целесообразно применять кумулянтные коэффициенты γ_4 и γ_6.

2. Кумулянтные коэффициенты

В общем случае кумулянтные коэффициенты γ_k определяются следующим образом [7]:

$$\gamma_k = \frac{\kappa_k}{\kappa_2^{\frac{k}{2}}}, \qquad (1)$$

где κ_k – кумулянты порядка k, однозначно связанные с центральными моментами μ_k. В частности, для симметричного распределения, которым является одновершинная двухкомпонентная гауссовская смесь

$$\begin{aligned} \kappa_1 &= m; \\ \kappa_2 &= \mu_2 = \sigma^2; \\ \kappa_3 &= \kappa_5 = 0; \\ \kappa_4 &= \mu_4 - 3\mu_2^2; \\ \kappa_6 &= \mu_6 - 15\mu_2\mu_4 + 30\mu_2^3, \end{aligned} \qquad (2)$$

где μ_k – центральные моменты.

Из формул (1) и (2) получаем выражения для кумулянтных коэффициентов γ_4 (коэффициент эксцесса) и γ_6:

$$\gamma_4 = \frac{\mu_4}{\mu_2^2} - 3,$$

$$\gamma_6 = \frac{\mu_6}{\mu_2^3} - 15\frac{\mu_4}{\mu_2^2} + 30.$$

На рис. 2 представлены коэффициенты γ_4 и γ_6 для ряда речевых сигналов, профильтрованных октавным фильтром со среднегеометрической частотой равной 125 Гц.

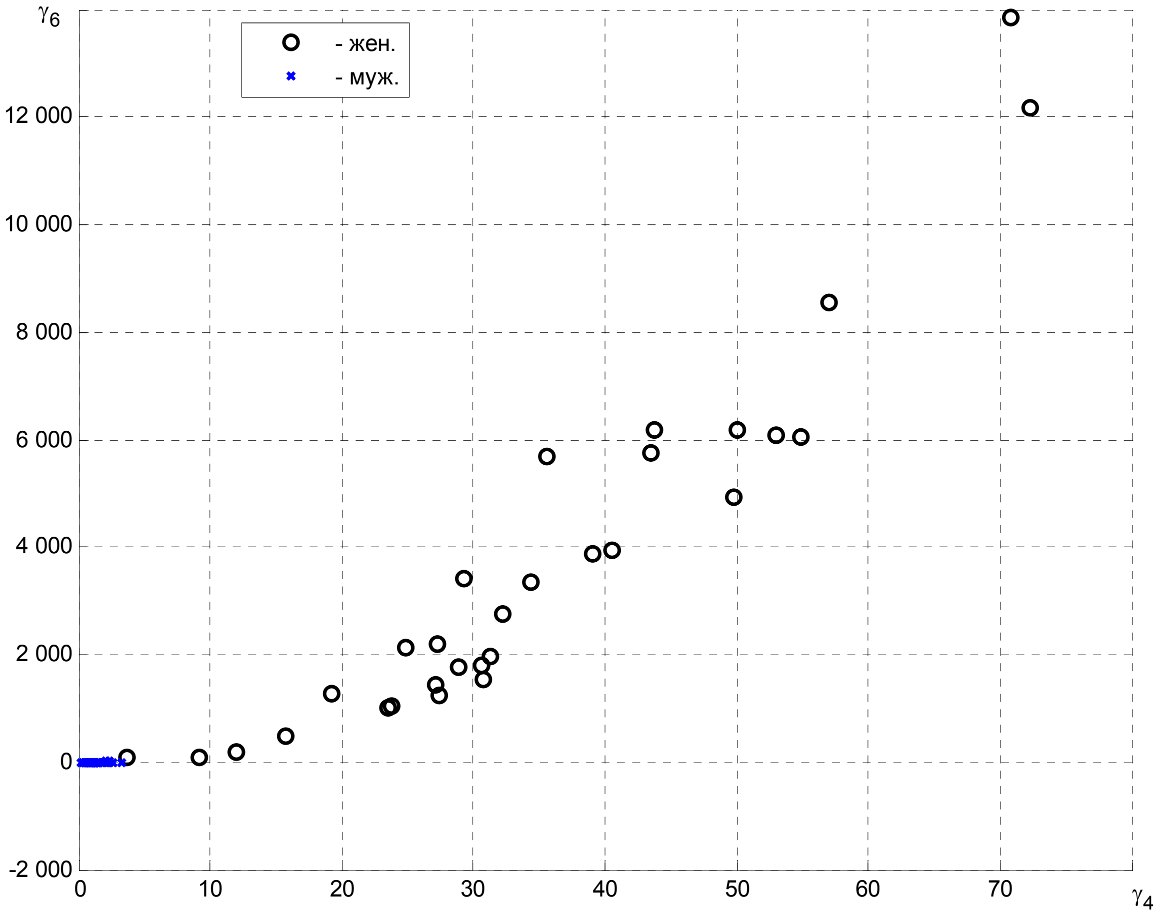

Рис. 2

Анализируя рис. 2 можно сделать вывод, что кумулянтные коэффициенты могут служить параметрами по которым можно определить пол диктора.

3. Характеристика речевой базы данных

Обучающая база данных сформирована на основе записей, сделанных в заглушенной комнате с усредненным временем реверберации 0,1 с. В записи участвовало 19 мужчин и 15 женщин. Для каждого диктора записано по два речевых фрагмента: один записан на украинском языке; второй – на русском. Таким образом база состоит из 68 сигналов длительностью 1 с каждый, записанных с частотой дискретизации 20050 Гц.

Контрольная (тестовая) база также состоит из 68 сигналов, записанных при тех же условиях что и сигналы обучающей базы.

4. Построение классификаторов

Логистическая регрессия

Для решения задач классификации часто применяется статистическая модель, именуемая логистической регрессией [8]. В частности, для бинарной классификации, при которой задача заключается в определении принадлежности объекта к одному из двух классов, классы можно обозначить как «0» и «1». В таком случае гипотеза будет иметь вид

$$h_{\boldsymbol\theta}(x) = g\left(\boldsymbol\theta^T x\right) = \frac{1}{1+e^{-\boldsymbol\theta^T x}}, \qquad (3)$$

где $g(z)$ – логистическая функция; $\boldsymbol\theta$ – вектор параметров, а процесс классификации будет заключается в вычислении функции $h_{\boldsymbol\theta}(x)$ для каждого классифицируемого объекта. Если значение функции (3) больше 0,5 то объект относится к классу «1», если меньше 0,5 – к классу «0».

Чтобы определить вектор параметров $\boldsymbol\theta$ необходимо иметь обучающую выборку, состоящую из вектора признаков x и вектора маркеров класса y.

Для получения вектора параметров $\boldsymbol\theta$ будем использовать метод максимального правдоподобия. В первую очередь запишем функцию правдоподобия

$$L(\boldsymbol\theta) = \prod_{i=1}^{m}\left(h_{\boldsymbol\theta}\left(x^{(i)}\right)\right)^{y^{(i)}}\left(1-h_{\boldsymbol\theta}\left(x^{(i)}\right)\right)^{1-y^{(i)}}.$$

Но максимизировать удобнее не функцию правдоподобия непосредственно, а логарифм функции правдоподобия:

$$\log L(\boldsymbol\theta) = \sum_{i=1}^{m} y^{(i)} \log h_{\boldsymbol\theta}\left(x^{(i)}\right) + \\ +\left(1-y^{(i)}\right)\log\left(1-h_{\boldsymbol\theta}\left(x^{(i)}\right)\right). \qquad (4)$$

Максимизацию функции (4) целесообразно выполнять применяя метод градиентного спуска, который позволяет посредством итераций получить параметры $\boldsymbol\theta$. На каждом итерационном шаге j-е значение вектора параметров $\boldsymbol\theta$ будет равно:

$$\theta_j := \theta_j + \alpha\left(y^{(i)} - h_{\boldsymbol\theta}\left(x^{(i)}\right)\right)x_j^{(i)}, \qquad (5)$$

где $:=$ – оператор, аналогичный по своему действию оператору $=$ в Matlab; α – шаг.

Применим выше описанный алгоритм к обучающей базе данных для получения параметров гипотезы (3). При этом классу «женщина» назначим индикатор «0», а классу мужчина – индикатор «1».

В результате получаем следующие значения параметров $\boldsymbol\theta$:

$$\theta_0 = 2{,}9454;$$
$$\theta_1 = 3{,}7294;$$
$$\theta_2 = -0{,}8335,$$

а гипотеза, соответственно, записывается следующим образом

$$h(\gamma_4,\gamma_6) = \frac{1}{1+\exp\{2{,}9454+3{,}7294\gamma_4-0{,}8335\gamma_6\}}. \qquad (6)$$

Построим на плоскости (γ_4,γ_6) границу, разделяющую оба класса. Для этого приравняем функцию (6) к 0,5 и выразим параметр γ_6 через параметр γ_4:

$$\gamma_6 = 1{,}1998(2{,}9454+3{,}7294\gamma_4). \qquad (7)$$

На рис. 3 представлены данные обучающей выборки и график границы классов (7) в разном масштабе.

Запишем гипотезу (6) в таком виде, чтобы получить классификатор, принимающий два значения «0» (для класса «женщина») и «1» (для класса «мужчина»).

Из формулы (7) и рис. 3 следует, что для всех записей женских голосов должно выполнятся условие

$$\gamma_6 > 1{,}1998(2{,}9454+3{,}7294\gamma_4), \qquad (8)$$

а для мужских –

$$\gamma_6 < 1{,}1998(2{,}9454+3{,}7294\gamma_4). \qquad (9)$$

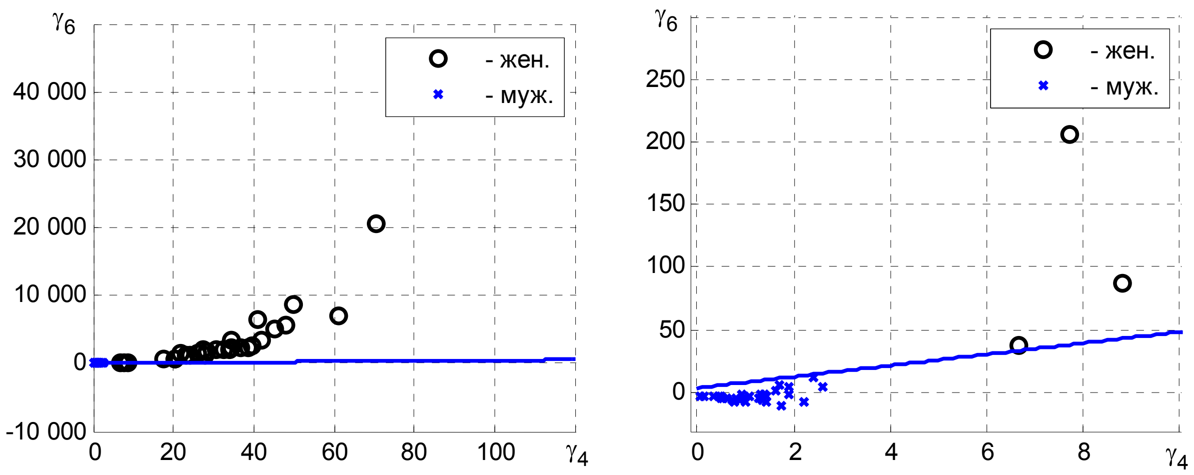

Рис. 3

Учитывая неравенства (8) и (9), запишем классификатор

$$\mathrm{H}(\gamma_4,\gamma_6)=\mathrm{sign}\left(1{,}1998(2{,}9454+3{,}7294\gamma_4)-\gamma_6\right). \quad (10)$$

Линейное разделение

Реализуем иной подход к построению классификатора. Как следует из рис. 2 исследуемые два класса являются линейно разделимыми. Это означает, что в заданном пространстве параметров можно построить гиперплоскость, которая разделит классы. Однако строить плоскость следует не произвольно, а таким образом, чтобы расстояние от объектов каждого класса до этой плоскости было максимально. Эта идея положена в основу целого набора алгоритмов, объеденных названием метод опорных векторов [8]. Следует отметить, что при построении разделяющей плоскости участвуют не все объекты классов, а лишь некоторые из них, которые и называются опорными.

В этой работе ограничимся наипростейшим случаем. Выберем из обоих классов по одному объекту, расстояние между которыми на плоскости $(\gamma_4;\gamma_6)$ минимально и проведем посредине между ними прямую

$$\gamma_6=k\gamma_4+b. \quad (1)$$

Отметим лишь, что данная прямая должна быть перпендикулярна прямой, соединяющей две выбранные точки.

Проведя несложные вычисления, получим параметры прямой (1):

$$k=-8{,}0208;$$

$$b=53{,}9592.$$

Таким образом граница между двумя классами описывается равенством

$$\gamma_6=-8{,}0208\gamma_4+53{,}9592, \quad (12)$$

а классификатор получим аналогично (10)

$$\mathrm{H}(\gamma_4,\gamma_6)=\mathrm{sign}(53{,}9592-\gamma_6-8{,}0208\gamma_4). \quad (13)$$

На рис. 4 в разных масштабах приведены данные обучающей выборки и график функции (12).

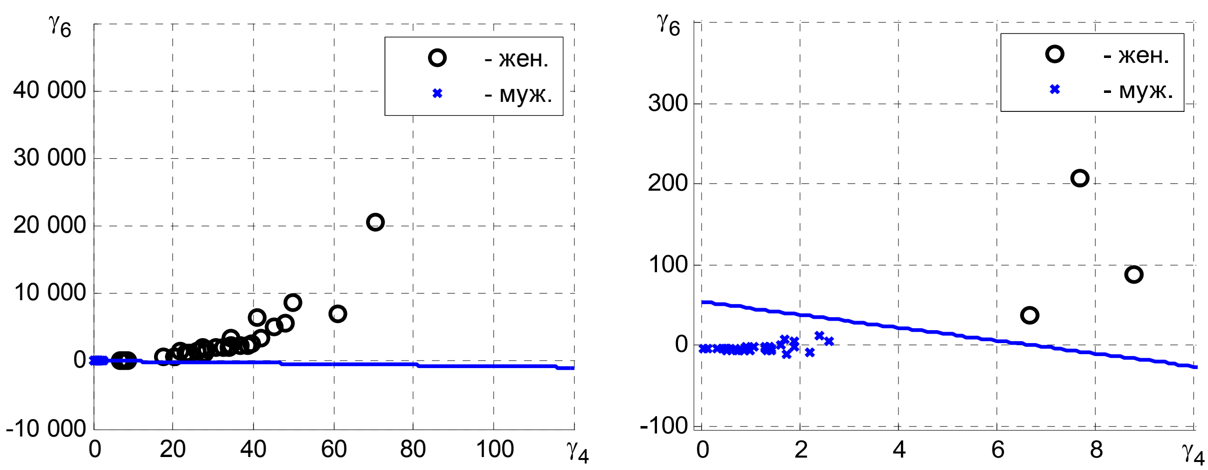

Рис. 4

5. Результаты классификации

В табл. 1 приведены проценты ошибок классификации сигналов тестовой базы с помощью классификаторов (10) и (13).

Таблица 1

Логистическая регрессия			
	Муж.	Жен.	Среднее
Ошибка, %	2,6	0	1,3
Линейное разделение			
	Муж.	Жен.	Среднее
Ошибка, %	0	0	0

Как следует из табл. 1 классификатор, полученный с помощью линейного разделения, дал лучший результат, то есть для данной тестовой базы все сигналы были правильно классифицированы.

Выводы

В работе показано, что анализируя параметры плотности вероятностей сигнала в октавной полосе частот со среднегеометрической частотой, раной 125 Гц, можно решать задачу классификации пола диктора.

В качестве классификационных признаков были выбраны кумулянтные коэффициенты γ_4 и γ_6, с их использованием были построены классификаторы.

Результаты тестирования показали, что классификатор, построенный с помощью линейного разделения дал лучший результат.

Литература

1. *Калюжный А.Я.*, Семенов В.Ю. Метод идентификации пола диктора на основе моделирования акустических параметров голоса гауссовыми смесями // Акустичний вісник. – 2009. – Т. 12, № 2. – С. 31–38.
2. *Scheme E.*, Castillo-Guerra E., Englehart K., Kizhanatham A. Practical Considerations for Real-Time Implementation of Speech-Based Gender Detection // Lecture notes in computer science. – 2006. – № 4225. – P. 426–436.
3. *Дидковский В.С.*, Продеус А.Н. Сопоставление формантных свойств украинской и русской речи // Электроника и связь. Тематический выпуск «Электроника и нанотехнологии». – 2009. – Ч.2. – С. 88–94.
4. *Величкин А.И.* Передача аналоговых сообщений по цифровым каналам связи. – М.: Радио и связь, 1983. – 240 с.
5. *Красильников А.И.*, Пилипенко К.П. Применение двухкомпонентной гауссовской смеси для идентификации одновершинных симметричных плотностей вероятностей // Электроника и связь. – 2008. – № 5(46). – С. 20–29.
6. *Красильников А.И.*, Пилипенко К.П. Одновершинная двухкомпонентная гауссовская смесь. Коэффициент эксцесса // Электроника и связь. – 2007. – № 2(37). – С. 32–38.
7. *Малахов А.Н.* Кумулянтный анализ случайных негауссовских процессов и их преобразований. – М.: Сов. радио, 1978. – 376 с.
8. *Friedman J.*, Hastie T., Tibshirani R. The Elements of Statistical Learning. – Springer. – 2008. – 739 p.

Национальный технический университет Украины «Киевский политехнический институт»

Поступила в редакцию 8 декабря 2011 г.